University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2009 *Paper* 250

A Machine-Learning Algorithm for Estimating and Ranking the Impact of Environmental Risk Factors in Exploratory Epidemiological Studies

Jessica G. Young* Alan E. Hubbard†

B Eskenazi‡ Nicholas P. Jewell**

*Harvard School of Public Health, jyoung@hsph.harvard.edu

†University of California, Berkeley, hubbard@stat.berkeley.edu

‡eskenazi@berkeley.edu

**Division of Biostatistics, School of Public Health, University of California, Berkeley, jewell@berkeley.edu

1 Introduction

Modern epidemiology faces many challenges in identifying real exposure effects on disease risks. This is particularly true in studies of environmental chemical exposures where population effects are likely small, measurement error is large, and *a priori* knowledge regarding the complex relationships between the many chemicals under study limited. Clearly, successful identification of real effects must start with proper study design, including limitation of measurement error and collection of sufficient data on confounders. However, even under the strong assumption of perfect study design, many exposure effects reported in the literature may not be real due to the approach to data analysis. Specifically, we consider three problems with 'typical' approaches to data analysis:

1. Failure to define a meaningful measure of effect;
2. Failure to account for the use of exploratory, data-adaptive methods with regard to inference;
3. Failure to account for multiple testing.

In this paper, we propose an algorithm designed to address each of these limitations in turn by combining recent advances in the causal inference and multiple-testing literature along with modifications to traditional non-parametric inference methods. Specifically, this algorithm does the following:

1. Estimates a population intervention model, for which we use a recently introduced parameter from the causal inference literature [Hubbard and van der Laan, 2005] naturally suited to environmental epidemiologic questions;
2. Provides marginal inference for this effect estimate based on a modified version of the conditional permutation test [Rosenbaum, 1984] to account for the presence of high-dimensional covariates using the propensity score;
3. Provides joint inference for multiple effect estimates using the Quantile Transformation Method [van der Laan and Hubbard, 2006].

We further describe an application of this algorithm to data collected from the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) project [Eskenazi et al., 2006]. This constitutes the first reported application of the estimation and inference methods referred to above to an environmental epidemiological data set. CHAMACOS is a longitudinal birth cohort study aimed at assessing the effects of pesticides and other environmental exposures on health outcomes in pregnant women and their children. The

CHAMACOS data set contains information on on birth outcomes among Latina mothers, neurobehavioral outcomes, hundreds of covariates and over one hundred exposure measurements representing at least 40 different chemicals. CHAMACOS is one of the first comprehensive studies of the impacts of chronic low-level pesticide exposure on human health, particularly in children.

In §2 we provide more background on the problems with 'typical' data analysis enumerated above and briefly describe how components of the proposed machine-learning algorithm address these problems. In §3 we define the data structure and parameters of interest. In §4 we describe the CHAMACOS data set in more detail. In §5 we describe estimation and inference for parameters of interest according to the components of the algorithm. In §6 we describe results of an application of this algorithm to the CHAMACOS data. In §7 we provide a discussion.

2 Background

1. Failure to define a meaningful measure of effect: As a simple abstract example, consider a study of the effect of some baseline level of a chemical exposure A on a continuously measured disease outcome Y. Further assume that a high-dimensional covariate vector W is observed containing sufficient data on confounders of the exposure effect. A typical approach to data analysis would likely involve regressing Y on a function of A and W. The likely reported exposure effect in this case would be the estimated coefficient on A in the postulated regression model. Assuming the regression model is correct, this coefficient represents the effect of A on Y, conditional on the covariates in W.

While this measure of effect has the advantage of being easily computed, it is often not very meaningful. Recall that the purpose of including W in our linear regression model is to adjust for confounding (remove bias). These variables are not included in the model because we are explicitly interested in the effect of A on Y conditional on *all* of these variables. The true parameter of interest is likely the marginal effect of A on Y, or, possibly this effect conditional on only a small subset of W, $V \in W$. For example, one might be interested in estimating the effect of exposure on the outcome separately in women and men.

Estimation of marginal structural models (MSM) introduced by Robins [1998] provides an alternative to traditional regression approaches. In particular, the parameters of MSMs, which model the distributions of counterfactual or potential outcomes under hypothetical exposure levels, represent marginal or conditional effects directly of interest. MSMs are estimated using inverse-probability weighting (IPW) or doubly-robust extensions (DR-IPW) [van der Laan and Robins, 2003]. The first component of the proposed algorithm discussed here consists of estimation of a recently proposed new class of models based on the MSM termed population intervention models [Hubbard and van der Laan, 2005]. These models

are also estimated using IPW or DR-IPW and are particularly relevant to population-based studies of risk factors.

2. Failure to account for the use of exploratory, data-adaptive methods with regard to inference: In both traditional regression approaches and estimation of MSMs, we are required to make modeling assumptions about nuisance parameters; that is, aspects of the true data-generating mechanism beyond assumptions about the exposure effect of interest. In traditional regression approaches such as that described above, we must make assumptions regarding the distribution of Y given A and W. In estimation of MSMs, we must make assumptions regarding either the distribution of Y given A and W or the distribution of A given W (often referred to as the 'treatment mechanism'). Data-adaptive approaches are ubiquitously used to estimate these nuisance parameters. In the case of traditional regression approaches, the use of such techniques is rarely reported. Even if they are reported, they are even less frequently accounted for in inference on estimated exposure effects. Alternatively, parametric assumptions are usually used to make inference that no longer hold when the form of nuisance parameters are selected data-adaptively.

The second component of the proposed algorithm attempts to address this problem through the use of a version of the conditional permutation test designed to provide more robust inference for the IPW and DR-IPW population intervention model estimators.

3. Failure to account for multiple testing: High-dimensional epidemiological data sets often involve testing multiple exposure effects on multiple outcomes. Despite this, most reports fail to account for multiple-testing in claims of statistical significance. This is mainly the result of the overly conservative nature of well-known multiple testing procedures (MTPs) such as the Bonferroni procedure.

The third component of the algorithm implements the Quantile Transformation Method [van der Laan and Hubbard, 2005], an MTP which appropriately adjusts for multiple testing (i.e. controls the appropriate type I error rate at some desired level α) and has been shown to be more powerful in simulation studies to alternative MTPs [Chen et al., 2007].

More detail regarding each of the components of the algorithm is provided in §5.

3 Data structure and parameters of interest

We assume we observe n i.i.d. copies of

$$O = \{W, A, Y\}.$$

We define W as a p-dimensional vector of covariates; $A = (A_1, \ldots, A_j, \ldots A_q)$ where A_j is the j^{th} exposure of interest; $Y = (Y_1, \ldots, Y_k, \ldots, Y_r)$ where Y_k is the k^{th} outcome of interest.

We assume a time-ordering of variables such that W precedes A which precedes Y. In order to define our parameter of interest, we view this observed data structure as a missing data structure, where the full, unobserved data structure consists of both the observed data and, for all j, k, the possibly unobserved counterfactual outcome $Y_{a_j,k}$, $a_j \in \mathcal{A}_j$. Specifically, $Y_{a_j,k}$ represents the k^{th} outcome an individual *would have* experienced had they, possibly contrary to fact, received level a_j for the j^{th} exposure of interest, for $\mathcal{A}_j$ the set of all possible levels of this exposure.

We make the following three identifying assumptions in order to link the observed and full data structures for all j, k:

1. Consistency assumption: $A_j = a_j \implies Y_{a_j,k} = Y_k$.
2. No unmeasured confounding (sequential randomization) assumption: $Y_{a_j,k} \coprod A_j | W$.
3. Experimental treatment assignment (ETA) assumption: $0 < Pr(A_j = a_j|W) < 1$ for all $a_j \in \mathcal{A}_j$.

Of these identifying assumptions, only violations of the ETA assumption may be empirically examined based on the observed data.

We define the parameter of interest, for all j, k, by:

$$\psi_{a_j,k} = E[Y_{a_j,k}] - E[Y_k]. \tag{1}$$

A population intervention model is a model for this parameter [Hubbard and van der Laan, 2005]. When $a_j = 0$ represents the level *unexposed*, $\psi_{0_j,k}$ may be interpreted as the effect of removing the j^{th} exposure on the mean of the k^{th} outcome in the target population (a measure akin to attributable risk).

Here we will consider only the marginal causal effect (1). However, as formalized in Hubbard and van der Laan [2005], (1) may be extended to models conditional on a subset of covariates $V \in W$. That is, we may define the alternative parameter:

$$\psi(V)_{a_j,k} = E[Y_{a_j,k}|V] - E[Y_k|V].$$

Here we interpret $\psi(V)_{0_j,k}$ as the effect of removing the j^{th} exposure on the mean of the k^{th} outcome in the target population within specific strata of V (e.g. amongst men or amongst women).

For simplicity, we will also only consider constant models for (1) or $\psi_{a_j,k} = \beta_{a_j,k}$. In general, however, (1) may be a function of a_j such that $\psi_{a_j,k} = m(a_j|\beta_{j,k})$ for some Euclidean parametrization $\beta_{j,k} \to m(a_j|\beta_{j,k})$ Hubbard and van der Laan [2005]. Allowing (1) to be a function of a_j can reduce variability of estimators when a_j has many categories or is measured continuously. This is at the expense of bias if this parametric assumption is incorrect.

4 CHAMACOS data description

Study participants were recruited amongst pregnant women initiating prenatal care at Natividad Medical Center, a county hospital in the city of Salinas, California, or at Clínica de Salud del Valle de Salinas in the Salinas Valley, California. The recruitment sites serve a majority of low-income individuals, with a large proportion working in agriculture. Eligible women were less than 20 weeks gestation, 18 years or older, Medi-cal eligible, fluent in English and/or Spanish, and planning to deliver at Natividad Medical Center. A total of 601 women were enrolled between October 1999 and October 2000. Of these, 536 continued in the study through delivery. Based on these 536 deliveries to 531 Latina women, the data set consists of information on 542 infants (536 live infants, 3 still-born infants, and 2 neonatal deaths). Chemical exposure measurements were taken at approximately 26 weeks gestation and again post-delivery (not all women had both samples) via maternal blood samples. For a more detailed description of data collection procedures, see Eskenazi et al. [2006], Fenster et al. [2007] and Chevrier et al. [2008].

A subset of the variables in the CHAMACOS data set is used for illustration purposes in this article. These consist of four birth outcomes (Y), 30 chemical exposures (A) and 13 covariates (W). Outcomes include birthweight (grams), gestational age (weeks), head circumference (cm), and length (cm). Exposures include 19 polychlorinated biphenyls (PCBs) (18, 28, 44, 49, 52, 66, 74, 99, 101, 118, 138, 146, 153, 156, 180, 183, 187, 194, 201) and 11 organochlorines (OCs). The 11 OCs include (table abbreviations in parentheses): β-hexachlorocyclohexane (BHC); Dieldrin (DIE); γ-hexachlorocyclohexane (GHC); hexachlorobenzene (HCB); Heptachlor epoxide (HPE); Mirex (MIR); *o,p´*-DDT (ODT); Oxychlordane (OXY); *p,p´*-DDE (PDE); *p,p´*-DDT (PDT); and *trans*-Nonachlor (TNA). The primary exposure levels used for all chemicals were those taken during pregnancy. However, if this level was missing and the post-delivery level was non-missing, the latter was used. All exposure measurements were measured in ng/g and lipid-adjusted. Values below the limit of detection (LOD) were imputed as LOD/2. Observations were assigned $A_j = 0$ if their observed level of the j^{th} chemical was in the bottom quartile of the empirical distribution of A_j.

Baseline covariates included: infant sex; pre-pregnancy BMI (underweight or normal, overweight, obese); marital status (single, married/living as married); poverty level (at or below poverty level, ≥200% poverty level); maternal education (≤6th grade, 7-12th grade, ≥high school graduate); parity (0, ≥1); number of years in the US (≤1, 2-5, 6-10, 11+); country of origin (US, Mexico or Other); gestational age at first prenatal visit (weeks); and maternal age at delivery (years). Poverty level was calculated by dividing household income by the number of people supported by that income and comparing this value to federal poverty thresholds [U.S. Census Bureau, 2000].

In estimating associations where the exposure of interest was a PCB, three OCs (*p,p´*-

DDE; *p,p´*-DDT; *o,p´*-DDT) were considered for any data-adaptive selection of models in addition to the baseline covariates W. For associations where this exposure was an OC, the sum of all 19 PCBs was considered in addition to the baseline covariates. Only non-missing PCBs were included in the sum for each observation.

The final analysis was limited to observations non-missing on all of these 13 covariates. This reduced the data set from 542 to 380 observations.

5 Methods

5.1 Estimation: IPW and DR-IPW estimation of the Population Intervention Model

Let $a_j = 0$ and $\psi_{a_j,k} \equiv \psi_{j,k}$. Using the estimating equation methodology of [van der Laan and Robins, 2003] the IPW and DR-IPW estimators of $\psi_{j,k}$ for the constant population intervention model $\psi_{j,k} = \beta_{j,k}$ are defined, respectively, as follows [Hubbard and van der Laan, 2005]:

$$\begin{aligned} \hat{\psi}_{j,k} &= \frac{1}{n}\sum_{i=1}^{n} D_i(j,k,\hat{g}_j) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=0)}{\hat{g}_j(0|W_i)} Y_i - \bar{Y} \end{aligned} \tag{2}$$

and

$$\begin{aligned} \hat{\psi}_{j,k} &= \frac{1}{n}\sum_{i=1}^{n} D_i(j,k,\hat{g}_j,\hat{Q}_{j,k}) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=0)}{\hat{g}_j(0|W_i)} Y_i - \bar{Y} \\ &- \frac{I(A_i=0)-\hat{g}_j(0|W_i)}{\hat{g}_j(0|W_i)} \hat{Q}_{j,k}(0,W_i), \end{aligned} \tag{3}$$

where $\hat{g}_j \equiv \hat{g}_j(0|W) \equiv \hat{Pr}(A_j = 0|W)$ and $\hat{Q}_{j,k} \equiv \hat{Q}_{j,k}(0,W) \equiv \hat{E}[Y_k|A_j = 0, W]$ are data-adaptively selected estimates of the nuisance parameters $g_j \equiv g_j(0|W) \equiv Pr(A_j = 0, W)$ and $Q_{j,k} \equiv Q_{j,k}(0,W) \equiv E[Y_k|A_j = 0, W]$, respectively. $\hat{\psi}_{j,k}$ as defined by (2) will be a consistent estimator of the true $\psi_{j,k}$ if the data-adaptively selected form of $\hat{g}_j(0|W)$ is correct. $\hat{\psi}_{j,k}$ as

defined by (3) will be a consistent estimator of the true $\psi_{j,k}$ if either the data-adaptively selected form of $\hat{g}_j(0|W)$ *or* the data-adaptively selected form of $\hat{Q}_{j,k}(0|W)$ is correct.

For the application of the machine learning algorithm, data-adaptively selected estimates of these nuisance parameters were obtained using the Deletion/Substitution/Addition (DSA) algorithm (Sinisi and van der Laan [2004]). The DSA algorithm is a data-adaptive modeling routine which uses cross-validation, based on the squared error (L2) loss function, to obtain a *best* model based on a set of candidate estimators. The space of candidate estimators is limited by three variables: the maximum allowable model size, the maximum order of interactions, and the maximum sum of powers on a single model term. In the estimation of both $g_j(0|W)$ and $Q_{j,k}(0,W)$, values for these limiting parameters were selected as 6, 2 and 3, respectively.

Estimates of $g_j(0|W)$ were truncated such that values were restricted above 0.1 in order to reduce variability associated with practical ETA violations. The possible implications of truncation, as well as the presence of practical ETA violations, on IPW estimates is discussed in §6.

5.2 Marginal inference: a modified conditional permutation test

Contrary to the estimators (2) and (3) above, assume that the correct forms of the nuisance parameters g_j and $Q_{j,k}$ are known *a priori* with corresponding maximum likelihood estimates $\hat{g}_j^*$ and $\hat{Q}_{j,k}^*$. Alternative IPW and DR-IPW estimators of $\psi_{j,k}$ to (2) and (3) are, in turn, defined as:

$$\hat{\psi}_{j,k}^* = \frac{1}{n}\sum_{i=1}^{n} D_i(j,k,\hat{g}_j^*) \tag{4}$$

and

$$\hat{\psi}_{j,k}^* = \frac{1}{n}\sum_{i=1}^{n} D_i(j,k,\hat{g}_j^*,\hat{Q}_{j,k}^*) \tag{5}$$

It has been shown [van der Laan and Robins, 2003, Hubbard and van der Laan, 2005] that, for both (4) and (5), $\sqrt{n}(\hat{\psi}_{j,k}^* - \psi_{j,k}^0) \longrightarrow N(0,\sigma_{j,k}^2)$ for $\psi_{j,k}^0$ the true population value of $\psi_{j,k}$. Conservative estimates of $\sigma_{j,k}^2$ can be obtained using

$$\hat{\sigma}_{j,k}^2 = \frac{\mathrm{var}(D(j,k,\hat{g}_j^*))}{n} \tag{6}$$

and

$$\hat{\sigma}_{j,k}^2 = \frac{\mathrm{var}(D(j,k,\hat{g}_j^*,\hat{Q}_{j,k}^*))}{n} \tag{7}$$

for (4) and (5), respectively, where $\text{var}(D(j,k,\hat{g}_j^*))$ and $\text{var}(D(j,k,\hat{g}_j^*,\hat{Q}_{j,k}^*))$ are obtained using the sample variance.

However, the forms of g_j and $Q_{j,k}$ are rarely known *a priori* and thus inference will almost always be based on the estimators (2) or (3). Here we propose an exact test of the null hypothesis that A_j and Y_k are independent based on the estimators (2) or (3) using a modification of the conditional permutation test.

First, we define how the conditional permutation test of this null hypothesis is implemented when W is of small dimension and consists of only covariates with few categories:

- Calculate the test statistic $T = \frac{\hat{\psi}_{j,k}}{\sqrt{\hat{\sigma}_{j,k}^2}}$ for $\hat{\psi}_{j,k}$ defined by (2) or (3) and, accordingly, $\hat{\sigma}_{j,k}^2$ defined by (6) or (7), using the original observed data.
- Randomly permute the values of A_j within each stratum of W and recalculate the test statistic based on this permutation.
- Repeat the previous step B times to obtain a B-length vector of test statistics $T_0 = (T_{01}, ..., T_{0B})$.
- To obtain a p-value (p), calculate the proportion of elements in T_0 for which the absolute value exceeds the absolute value of the original test statistic calculated from the observed data.

T_0 is referred to as the conditional permutation distribution of T. The conditional permutation test rejects the null hypothesis of independence between A_j and Y_k given W if $p < \alpha$ for some pre-specified level α (e.g. $\alpha = 0.05$).

Note that this testing procedure will still retain correct type I error control at level α even when the test statistic is defined using poor estimates of $\sigma_{j,k}^2$ (which, as noted above, may be the case when estimators of the exposure effect are based on data-adaptively selected estimates of nuisance parameters). Testing procedures using the conditional permutation distribution as the assumed null will still retain correct type I error control when the standard error is over or underestimated due to the fact that the test statistic can be defined as any algorithm of the data.

In practice however, testing using the conditional permutation test may be infeasible. Specifically, permutation within W becomes difficult when W is high-dimensional or even has one element with continuous values. Here we present an *ad hoc* modification of the conditional permutation test for more general W which uses $\hat{g}_j$, the estimated treatment mechanism or propensity score, when g_j is assumed to follow a logistic model. This modified approach is implemented as the second step of the proposed machine-learning algorithm. Specifically,

- Obtain an estimate of the propensity score, $\hat{g}_j(0|W)$, for each observation using standard logistic regression.
- Order the data by these estimated probabilities, $\hat{g}_j(0|W)$.
- Group observations so that, within each group, the minimum number with $A_j = 0$ or $A_j = 1$ is M. This grouping constitutes a new categorical variable W^*.
- Follow the steps above for performing the conditional permutation test, permuting within strata of W^* in place of W.

Use of this version of the conditional permutation test assumes that each category of W^* will contain individuals with comparable values on the covariates W. The method of grouping we propose is one of a variety of *ad hoc* procedures which have been suggested for obtaining these categories based on the propensity score, including that originally proposed by Rosenbaum [1984] based on a backtrack algorithm. Simulation studies have suggested that defining W^* using our approach successfully controls the type I error rate at the desired level α.

In the application to CHAMACOS, the conditional permutation distribution for each test was based on $B = 5,000$ permutations, with categories for W^* defined by $M = 2$. As discussed extensively above, $\hat{g}_j(0|W)$ generally represents a data-adaptively selected estimate of the true propensity score. In the application of these methods to the CHAMACOS data set, we found that programs took an unreasonable amount of time to run when the form of $\hat{g}_j(0|W)$ was reselected data-adaptively within each permutation iteration. To address this issue, forms for $\hat{g}_j$ selected using the DSA on the original data were reused, with model coefficients re-estimated using maximum likelihood within each permutation.

5.3 Joint inference: Quantile Transformation Method

The previous section describes marginal inference based on estimates of the effect of a single exposure A_j on a single outcome Y_k. Usually we will be interested in testing multiple exposure-outcome associations. If we are testing hypotheses regarding all exposure-outcome combinations, we will have a total of $m = q \times r$ tests. As described in §4, the CHAMACOS data set consists of 30 chemical exposures and 4 outcomes, resulting in a total of 120 tests.

While in environmental epidemiology the value of m is generally substantially greater than one, reported approaches to inference usually ignore the multiple testing problem. In particular, these reports fail to correctly define the false positive or type I error rate in terms of the *total* number of false positives, V_n. Note that V_n is a random variable which can take on the values zero or one in the case where $m = 1$ and any value between zero and the total number of true null hypotheses amongst the m tests in the case where $m > 1$. There

are various forms for the type I error rate, depending on the type and stringency of control desired by the investigator. For the purposes of this discussion and in the application below, we focus on controlling the family-wise error rate (FWER) or $P(V_n > 0)$ at α.

Failure to correctly define the type I error rate appropriately in most reported epidemiologic investigations is likely due to the fact that the well-known and easily implemented Bonferroni procedure is overly conservative. Specifically, it requires the (unlikely true) assumption that all m tests are independent, rejecting the null hypothesis for $p < \frac{\alpha}{m}$. Intuitively, one can imagine that if all m tests are perfectly correlated, we should divide by one in place of m. Multiple testing procedures (MTPs) which, in addition to correctly controlling the type I error rate below α, further maximize power through use of information on the joint dependence structure of the test statistics are preferable.

The third step of the algorithm implements one such approach, referred to as the Quantile Transformation Method (van der Laan and Hubbard [2005]). The Quantile Transformation Method is a resampling-based method which essentially incorporates the desirable characteristics of currently available MTPs, including the use of information on the dependence structure of the test statistics. This approach is an extension of a resampling-based method, originally proposed by Pollard and van der Laan [2003] and further developed by Dudoit et al. [2004], which creates an appropriate joint null distribution using the bootstrap. The observed m test statistics are then compared to this estimated joint null to obtain p-values. P-values obtained via some MTP are generally referred to as *adjusted* and those obtained otherwise as *raw* or *marginal*.

To implement the Quantile Transformation method:

- Sample with replacement (or bootstrap) the observed data (W, A, Y) amongst the n observations.
- Calculate the m-length vector of test statistics based on this new sample.
- Repeat these first two steps B_2 times to obtain a $m \times B_2$ matrix of test statistics, $T^{\#}$, where $T^{\#}_{l,b}$ represents the test statistic obtained from the b^{th} bootstrap sample for the l^{th} test, $l = 1, \ldots, m$, $b = 1, ..., B_2$.
- Using the l^{th} row of $T^{\#}$, calculate the empirical bootstrap distribution Q_{nl} for the l^{th} test statistic T_l, where $Q_{nl}(t) = P_n(T_l < t)$. This results in a $m \times B_2$ matrix of estimated probabilities, Q_n.
- Apply the quantile-function (or inverse probability function) $Q^{-1}_{0l}(x)$ to the l^{th} row of Q_n, defining Q^{-1}_{0l} in terms of the assumed null distribution for the l^{th} test statistic. This maps Q_n to a new $m \times B_2$ matrix, Q^{-1}_0, representing a joint null distribution for the observed m test statistics. For example, for $x = 0.5$, $Q^{-1}_{0l}(x)$ is the median of the null distribution for the l^{th} observed test statistic. Note that Q^{-1}_{0l} can be the

inverse probability function for any desired marginal null distribution for the l^{th} test statistic. In our case, we define the marginal null distribution in terms of the modified conditional permutation distribution described in §5.2.

Once the joint null distribution, Q_0^{-1} for the m test statistics is obtained, various MTPs can be applied to obtain an adjusted p-value. In our application to CHAMACOS we apply the single-step minP approach, which converts the matrix Q_0^{-1} to p-values based on the distribution of each row. The adjusted p-value is then obtained by comparing the l^{th} raw p-value to the distribution of the minimum from each column. The FWER is controlled at α by rejecting the null hypothesis for a given test when the adjusted p-value is less than α. For an extensive review of various MTPs, including minP and maxT, see Dudoit et al. [2004], van der Laan et al. [2004b], and van der Laan et al. [2004a].

In our data application, the joint null distribution of the test statistics was estimated using $B_2 = 5,000$ bootstrap samples. As in the case of the modified conditional permutation tests, we found unreasonable computing times when the forms of $\hat{g}_j$ and $\hat{Q}_{j,k}$ were re-selected using the DSA within each bootstrap iteration. Again, here, the forms of $\hat{g}_j$ and $\hat{Q}_{j,k}$ chosen by the DSA using the original sample data were re-used within each bootstrap iteration with model coefficients re-estimated using maximum likelihood based on the bootstrap sample.

6 Results

Tables 1 through 4 present results based on estimates of (1) unadjusted for W. The form of the unadjusted estimator is identical to that of the IPW estimator (2) with the weights $\hat{g}_j$ calculated as the proportion of observations with $A_j = 0$ (this is simply the difference between the mean of the outcome amongst the baseline exposure group or $A = 0$ and the overall mean). All p-values (both raw and adjusted) in Tables 1 through 4 are obtained from the simple permutation distribution; that is, only the A_j's are permuted as there are no variables in W considered. Tables 5 through 8 and 9 through 12 present results based on the IPW and DR-IPW estimates of (1), respectively. For comparison purposes, Tables 5 through 12 present p-values based on both the conditional permutation distribution and the standard Normal. Adjusted p-values are presented based on both the Bonferroni and Quantile Transformation Methods. Results are separated by outcome for ease of presentation, however, adjusted p-values take into account all 120 tests.

Out of the 120 tests of association, the algorithm found only one significant association defining $\alpha = 0.1$ between HPE ($j = 24$) and head circumference ($k = 3$) with the DR-IPW estimator $\hat{\psi}_{24,3} = -0.507$ and an adjusted p-value of 0.079 obtained using the Quantile Tranformation Method with marginal distribution defined by the modified conditional permutation distribution. This suggests, assuming our identifying assumptions hold,

that infant head circumference would decrease 0.507 cm on average were all maternal HCB levels changed to the bottom quartile of the observed distribution compared to the mean of observed maternal HCB levels.

Note that results based on the unadjusted estimate of (1) and the IPW adjusted estimate should be identical when the model selected by the DSA for $\hat{g}_j(0|W)$ is constant (i.e. no variables in W predict exposure). Despite this, marginal p-values based on the unadjusted estimator in Tables 1 through 4 differ slightly from marginal p-values based on the IPW estimator reported in Tables 5 through 8, even in the case where $\hat{g}_j(0|W)$ is constant. This is because the latter are obtained by conditionally permuting A_j within levels of W^* whereas the former are obtained by simply permuting A_j marginally. In this case, the unadjusted estimator is to be preferred. We recalculated the Quantile Transformation Method adjusted p-values replacing the conditional permutation distributions of the IPW estimator with the simple permutation distribution of the unadjusted estimator in cases where the model selected by the DSA for $\hat{g}_j(0|W)$ was constant and resulting adjusted p-values were essentially unchanged.

Again, the consistency of our IPW and DR-IPW estimates of (1) relies on the ETA assumption. Our use of truncated weights is an attempt to reduce variability in the presence of possible practical ETA violations at a cost to bias resulting from misspecification of the treatment mechanism (estimates in Tables 5 through 12 are starred where truncation is used). There are methods to examine the bias due to the ETA assumption [Wang et al., 2006], as well as other parameters one could estimate that are less like to suffer from such a bias [van der Laan and Petersen, 2007]. For now, we take this relative simple, ad hoc approach, noting that other more detailed diagnostics and alternative parameter estimates are available.

To illustrate differences between the standard Normal and the modified conditional permutation distribution, Figure 1 overlays the cumulative distribution functions of these two possible distributions for the test statistic associating head circumference and PCB 18 based on the DR-IPW estimator. It is clear from this figure that the modified conditional permutation distribution is similar to that of the standard Normal but with variance less than one. Table 14 presents standard errors for this same estimator based on both the influence curve as defined in equation (7) and based on 5000 bootstrap samples. The standard error estimate based on the bootstrap is smaller than that based on the influence curve, suggesting the latter is overly conservative. Similarly, we see from Table 14 that the bootstrap estimate of the variance of the test statistic (also based on 5000 samples) is smaller than one.

Table 15 provides an overall summary of the bootstrapped estimates of the variance of the test statistic for all 120 tests based on both the IPW and DR-IPW estimators. In almost all cases we see that the standard error estimate based on the influence curve using (6) and (7) is likely overly conservative. Specifically, for all but one test statistic based on the DR estimator, the bootstrapped estimate of the variance is less than one.

Table 1: Unadjusted estimates for associations between birthweight and each exposure (exp), number of observations nonmissing on exposure and birthweight (N), standard errors (SE), test statistics (T), raw, Bonferroni (Bon) adjusted, and Quantile Transformation Method (QTM) adjusted p-values from the simple permutation distribution.

exp	N	Estimate	SE	T	Raw	Bon	QTM
18	370	13.005	317.976	0.041	0.767	1	1
28	380	28.418	316.007	0.090	0.523	1	1
44	314	-0.170	344.926	0	0.996	1	1
49	331	7.162	337.755	0.021	0.883	1	1
52	346	4.949	328.513	0.015	0.914	1	1
66	369	-41.358	316.033	-0.131	0.364	1	1
74	361	31.024	326.411	0.095	0.499	1	1
99	348	1.842	327.128	0.006	0.968	1	1
101	320	-30.840	339.483	-0.091	0.527	1	1
118	357	-24.156	323.043	-0.075	0.592	1	1
138	349	13.760	331.318	0.042	0.771	1	1
146	332	51.385	339.956	0.151	0.289	1	1
153	358	38.775	324.686	0.119	0.410	1	1
156	363	12.080	321.566	0.038	0.790	1	1
180	299	78.942	360.732	0.219	0.121	1	1
183	348	-13.530	327.068	-0.041	0.769	1	1
187	301	5.280	349.869	0.015	0.920	1	1
194	352	23.184	325.597	0.071	0.617	1	1
201	364	86.716	327.164	0.265	0.062	1	0.994
BHC	378	-30.137	312.673	-0.096	0.485	1	1
DIE	359	-55.315	317.487	-0.174	0.210	1	1
GHC	374	-41.242	310.737	-0.133	0.344	1	1
HCB	380	10.597	314.626	0.034	0.804	1	1
HPE	363	-99.235	311.674	-0.318	0.024	1	0.874
MIR	379	22.162	315.213	0.070	0.616	1	1
ODT	380	-48.624	310.020	-0.157	0.277	1	1
OXY	358	-21.733	319.736	-0.068	0.633	1	1
PDE	380	2.650	314.704	0.008	0.947	1	1
PDT	380	-41.688	311.300	-0.134	0.338	1	1
TNA	380	1.901	313.407	0.006	0.965	1	1

Table 2: Unadjusted estimates for associations between gestational age and each exposure (exp), number of observations nonmissing on exposure and gestational age (N), standard errors (SE), test statistics (T), raw, Bonferroni (Bon) adjusted, and Quantile Transformation Method (QTM) adjusted p-values from the simple permutation distribution.

exp	N	Estimate	SE	T	Raw	Bon	QTM
18	370	0.138	3.506	0.039	0.318	1	1
28	380	0.013	3.460	0.004	0.909	1	1
44	314	-0.047	3.800	-0.012	0.753	1	1
49	331	0.046	3.707	0.012	0.750	1	1
52	346	0.102	3.618	0.028	0.477	1	1
66	369	-0.212	3.497	-0.061	0.129	1	1
74	361	-0.009	3.552	-0.002	0.945	1	1
99	348	0.072	3.624	0.020	0.626	1	1
101	320	-0.034	3.766	-0.009	0.814	1	1
118	357	-0.093	3.570	-0.026	0.510	1	1
138	349	0.094	3.653	0.026	0.530	1	1
146	332	0.286	3.729	0.077	0.054	1	0.989
153	358	0.123	3.563	0.035	0.393	1	1
156	363	0.127	3.544	0.036	0.358	1	1
180	299	0.084	3.900	0.022	0.586	1	1
183	348	0.333	3.646	0.091	0.019	1	0.808
187	301	0.196	3.879	0.051	0.205	1	1
194	352	0.134	3.606	0.037	0.358	1	1
201	364	0.165	3.551	0.046	0.237	1	1
BHC	378	-0.086	3.451	-0.025	0.535	1	1
DIE	359	0.078	3.559	0.022	0.585	1	1
GHC	374	-0.031	3.472	-0.009	0.817	1	1
HCB	380	0.161	3.474	0.046	0.235	1	1
HPE	363	0.055	3.539	0.016	0.680	1	1
MIR	379	-0.050	3.454	-0.014	0.712	1	1
ODT	380	-0.050	3.457	-0.014	0.727	1	1
OXY	358	0.162	3.564	0.045	0.251	1	1
PDE	380	0.013	3.462	0.004	0.935	1	1
PDT	380	-0.039	3.458	-0.011	0.785	1	1
TNA	380	0.118	3.470	0.034	0.379	1	1

Table 3: Unadjusted estimates for associations between head circumference and each exposure (exp), number of observations nonmissing on exposure and head circumference (N), standard errors (SE), test statistics (T), raw, Bonferroni (Bon) adjusted, and Quantile Transformation Method (QTM) adjusted p-values from the simple permutation distribution.

exp	N	Estimate	SE	T	Raw	Bon	QTM
18	360	0.119	3.137	0.038	0.381	1	1
28	370	0.157	3.086	0.051	0.242	1	1
44	306	0.169	3.424	0.049	0.250	1	1
49	323	0.106	3.312	0.032	0.461	1	1
52	336	0.104	3.274	0.032	0.473	1	1
66	359	-0.014	3.125	-0.004	0.921	1	1
74	351	0.045	3.163	0.014	0.754	1	1
99	340	-0.159	3.229	-0.049	0.256	1	1
101	311	0.089	3.396	0.026	0.551	1	1
118	347	-0.112	3.168	-0.035	0.420	1	1
138	340	-0.138	3.228	-0.043	0.324	1	1
146	324	-0.077	3.283	-0.023	0.597	1	1
153	348	-0.094	3.171	-0.030	0.497	1	1
156	353	-0.200	3.076	-0.065	0.124	1	1
180	290	0.109	3.455	0.031	0.465	1	1
183	339	-0.253	3.166	-0.080	0.064	1	0.995
187	292	-0.187	3.354	-0.056	0.190	1	1
194	343	-0.056	3.141	-0.018	0.683	1	1
201	354	-0.166	3.123	-0.053	0.213	1	1
BHC	368	-0.152	3.078	-0.050	0.251	1	1
DIE	350	-0.254	3.133	-0.081	0.056	1	0.991
GHC	364	-0.122	3.098	-0.039	0.351	1	1
HCB	370	-0.197	3.077	-0.064	0.136	1	1
HPE	353	-0.483	3.094	-0.156	0	0	0
MIR	369	-0.038	3.024	-0.013	0.768	1	1
ODT	370	-0.236	3.073	-0.077	0.075	1	0.997
OXY	349	-0.259	3.108	-0.083	0.054	1	0.989
PDE	370	-0.021	3.115	-0.007	0.877	1	1
PDT	370	-0.149	3.081	-0.049	0.257	1	1
TNA	370	-0.157	3.036	-0.052	0.226	1	1

Table 4: Unadjusted estimates for associations between length and each exposure (exp), number of observations nonmissing on exposure and length (N), standard errors (SE), test statistics (T), raw, Bonferroni (Bon) adjusted, and Quantile Transformation Method (QTM) adjusted p-values from the simple permutation distribution.

exp	N	Estimate	SE	T	Raw	Bon	QTM
18	366	-0.351	4.501	-0.078	0.122	1	1
28	376	-0.172	4.472	-0.038	0.450	1	1
44	312	-0.270	4.896	-0.055	0.278	1	1
49	328	-0.391	4.766	-0.082	0.106	1	1
52	342	-0.211	4.668	-0.045	0.372	1	1
66	365	-0.228	4.543	-0.050	0.313	1	1
74	357	0.111	4.628	0.024	0.630	1	1
99	344	0.114	4.733	0.024	0.633	1	1
101	316	-0.380	4.857	-0.078	0.117	1	1
118	354	0.061	4.687	0.013	0.785	1	1
138	345	0.269	4.786	0.056	0.272	1	1
146	328	0.276	4.873	0.057	0.260	1	1
153	354	0.218	4.658	0.047	0.360	1	1
156	359	0.092	4.598	0.020	0.694	1	1
180	296	0.538	5.153	0.104	0.037	1	0.956
183	345	0.188	4.714	0.040	0.423	1	1
187	297	0.295	5.035	0.059	0.247	1	1
194	348	0.071	4.696	0.015	0.774	1	1
201	360	0.397	4.622	0.086	0.082	1	0.999
BHC	374	0.052	4.555	0.011	0.826	1	1
DIE	355	-0.212	4.626	-0.046	0.353	1	1
GHC	370	-0.138	4.497	-0.031	0.543	1	1
HCB	376	-0.114	4.478	-0.025	0.609	1	1
HPE	359	-0.313	4.592	-0.068	0.166	1	1
MIR	375	0.145	4.500	0.032	0.518	1	1
ODT	376	0.189	4.539	0.042	0.411	1	1
OXY	354	-0.083	4.599	-0.018	0.724	1	1
PDE	376	0.220	4.605	0.048	0.347	1	1
PDT	376	0.198	4.571	0.043	0.397	1	1
TNA	376	-0.011	4.488	-0.002	0.965	1	1

Table 5: IPW estimates for associations between birthweight and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	13.005	317.976	0.041	0.777	0.967	1	1	1	1
28	29.969	333.390	0.090	0.587	0.928	1	1	1	1
44	-0.170	344.926	0	0.999	1	1	1	1	1
49	7.162	337.755	0.021	0.951	0.983	1	1	1	1
52	4.949	328.513	0.015	0.930	0.988	1	1	1	1
66	-59.061	324.394	-0.182	0.153	0.856	1	1	1	1
74*	-143.079	340.887	-0.420	0.816	0.675	1	1	1	1
99*	-342.743	332.709	-1.030	0.281	0.303	1	1	1	1
101	-45.363	350.850	-0.129	0.234	0.897	1	1	1	1
118*	-25.186	372.911	-0.068	0.651	0.946	1	1	1	1
138*	-175.488	378.153	-0.464	0.485	0.643	1	1	1	1
146*	-9.010	390.706	-0.023	0.903	0.982	1	1	1	1
153*	-257.281	370.378	-0.695	0.775	0.487	1	1	1	1
156*	-154.631	357.369	-0.433	0.276	0.665	1	1	1	1
180*	-126.013	431.278	-0.292	0.949	0.770	1	1	1	1
183*	-293.159	363.023	-0.808	0.219	0.419	1	1	1	1
187*	-498.049	373.393	-1.334	0.180	0.182	1	1	1	1
194*	-426.147	338.862	-1.258	0.542	0.209	1	1	1	1
201*	-259.991	341.214	-0.762	0.767	0.446	1	1	1	1
BHC*	-458.087	361.062	-1.269	0.142	0.205	1	1	1	1
DIE*	-84.134	378.626	-0.222	0.669	0.824	1	1	1	1
GHC	-5.456	349.569	-0.016	0.919	0.988	1	1	1	1
HCB*	-726.817	320.020	-2.271	0.240	0.023	1	1	1	0.905
HPE*	-205.299	348.859	-0.588	0.054	0.556	1	1	0.990	1
MIR*	-17.829	342.523	-0.052	0.888	0.958	1	1	1	1
ODT	-26.123	341.175	-0.077	0.541	0.939	1	1	1	1
OXY*	-337.933	362.684	-0.932	0.150	0.351	1	1	1	1
PDE	52.237	348.010	0.150	0.339	0.881	1	1	1	1
PDT*	-35.304	387.776	-0.091	0.722	0.927	1	1	1	1
TNA*	-337.025	359.918	-0.936	0.569	0.349	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 6: IPW estimates for associations between gestational age and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	0.138	3.506	0.039	0.150	0.969	1	1	1	1
28	0.188	3.682	0.051	0.846	0.959	1	1	1	1
44	-0.047	3.800	-0.012	0.928	0.990	1	1	1	1
49	0.046	3.707	0.012	0.670	0.990	1	1	1	1
52	0.102	3.618	0.028	0.370	0.978	1	1	1	1
66	-0.186	3.634	-0.051	0.073	0.959	1	1	0.998	1
74*	-2.092	3.658	-0.572	0.615	0.567	1	1	1	1
99*	-3.717	3.733	-0.996	0.510	0.319	1	1	1	1
101	0.028	3.944	0.007	0.812	0.994	1	1	1	1
118*	-0.295	4.078	-0.072	0.425	0.942	1	1	1	1
138*	-2.054	4.183	-0.491	0.360	0.623	1	1	1	1
146*	-0.120	4.372	-0.028	0.833	0.978	1	1	1	1
153*	-2.937	4.131	-0.711	0.707	0.477	1	1	1	1
156*	-1.581	3.993	-0.396	0.439	0.692	1	1	1	1
180*	-2.429	4.587	-0.529	0.782	0.596	1	1	1	1
183*	-2.481	4.117	-0.602	0.425	0.547	1	1	1	1
187*	-5.280	4.191	-1.260	0.210	0.208	1	1	1	1
194*	-5.068	3.721	-1.362	0.532	0.173	1	1	1	1
201*	-3.408	3.761	-0.906	0.664	0.365	1	1	1	1
BHC*	-4.671	4.124	-1.133	0.516	0.257	1	1	1	1
DIE*	-0.721	4.158	-0.173	0.709	0.862	1	1	1	1
GHC	0.409	3.912	0.104	0.341	0.917	1	1	1	1
HCB*	-7.587	3.675	-2.064	0.710	0.039	1	1	1	0.980
HPE*	-1.219	3.991	-0.306	0.664	0.760	1	1	1	1
MIR*	-0.666	3.708	-0.180	0.978	0.857	1	1	1	1
ODT	0.012	3.788	0.003	0.935	0.998	1	1	1	1
OXY*	-3.368	4.059	-0.830	0.156	0.407	1	1	1	1
PDE	0.096	3.778	0.025	0.515	0.980	1	1	1	1
PDT*	-0.646	4.224	-0.153	0.915	0.878	1	1	1	1
TNA*	-3.424	4.056	-0.844	0.833	0.399	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 7: IPW estimates for associations between head circumference and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	0.119	3.137	0.038	0.161	0.970	1	1	1	1
28	0.157	3.086	0.051	0.255	0.959	1	1	1	1
44	0.169	3.424	0.049	0.243	0.961	1	1	1	1
49	0.106	3.312	0.032	0.262	0.974	1	1	1	1
52	0.104	3.274	0.032	0.337	0.975	1	1	1	1
66	-0.042	3.257	-0.013	0.811	0.990	1	1	1	1
74*	-1.651	3.316	-0.498	0.756	0.619	1	1	1	1
99*	-3.274	3.329	-0.983	0.285	0.325	1	1	1	1
101	0.084	3.583	0.023	0.637	0.981	1	1	1	1
118*	-0.422	3.510	-0.120	0.610	0.904	1	1	1	1
138*	-4.885	3.520	-1.388	0.584	0.165	1	1	1	1
146*	-0.668	3.816	-0.175	0.041	0.861	1	1	0.975	1
153*	-5.498	3.431	-1.603	0.348	0.109	1	1	1	1
156*	-2.340	3.491	-0.670	0.481	0.503	1	1	1	1
180*	-3.517	4.025	-0.874	0.818	0.382	1	1	1	1
183*	-2.478	3.613	-0.686	0.435	0.493	1	1	1	1
187*	-4.704	3.692	-1.274	0.732	0.203	1	1	1	1
194*	-5.525	3.255	-1.697	0.631	0.090	1	1	1	1
201*	-2.358	3.472	-0.679	0.451	0.497	1	1	1	1
BHC*	-4.190	3.501	-1.197	0.778	0.231	1	1	1	1
DIE*	-1.811	3.645	-0.497	0.324	0.619	1	1	1	1
GHC	-0.053	3.229	-0.016	0.721	0.987	1	1	1	1
HCB*	-6.653	3.248	-2.048	0.379	0.041	1	1	1	0.983
HPE*	-1.560	3.749	-0.416	0.055	0.677	1	1	0.992	1
MIR*	-0.773	3.399	-0.228	0.718	0.820	1	1	1	1
ODT	-0.183	3.304	-0.055	0.203	0.956	1	1	1	1
OXY*	-3.264	3.512	-0.930	0.099	0.353	1	1	1	1
PDE*	-1.778	3.419	-0.520	0.828	0.603	1	1	1	1
PDT	-0.037	3.559	-0.010	0.874	0.992	1	1	1	1
TNA*	-3.004	3.601	-0.834	0.711	0.404	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 8: IPW estimates for associations between length and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	-0.351	4.501	-0.078	0.496	0.938	1	1	1	1
28	-0.197	4.603	-0.043	0.707	0.966	1	1	1	1
44	-0.270	4.896	-0.055	0.476	0.956	1	1	1	1
49	-0.391	4.766	-0.082	0.250	0.935	1	1	1	1
52	-0.211	4.668	-0.045	0.470	0.964	1	1	1	1
66*	0.121	5.087	0.024	0.817	0.981	1	1	1	1
74*	-2.454	4.809	-0.510	0.731	0.610	1	1	1	1
99*	-4.685	4.886	-0.959	0.436	0.338	1	1	1	1
101	-0.412	5.238	-0.079	0.752	0.937	1	1	1	1
118*	-0.889	5.347	-0.166	0.296	0.868	1	1	1	1
138*	-6.634	5.260	-1.261	0.704	0.207	1	1	1	1
146*	0.088	5.871	0.015	0.885	0.988	1	1	1	1
153*	-3.468	5.419	-0.640	0.767	0.522	1	1	1	1
156*	-2.023	5.188	-0.390	0.359	0.697	1	1	1	1
180*	-5.628	5.834	-0.965	0.842	0.335	1	1	1	1
183*	-2.326	5.310	-0.438	0.789	0.661	1	1	1	1
187*	-3.795	5.265	-0.721	0.649	0.471	1	1	1	1
194*	-7.955	4.888	-1.627	0.443	0.104	1	1	1	1
201*	-0.540	5.088	-0.106	0.675	0.916	1	1	1	1
BHC	-0.010	5.798	-0.002	0.993	0.999	1	1	1	1
DIE	-0.264	5.126	-0.052	0.298	0.959	1	1	1	1
GHC	0.427	4.838	0.088	0.331	0.930	1	1	1	1
HCB*	-9.895	4.766	-2.076	0.382	0.038	1	1	1	0.978
HPE*	-3.558	5.432	-0.655	0.268	0.512	1	1	1	1
MIR*	-0.424	4.903	-0.087	0.971	0.931	1	1	1	1
ODT	0.315	4.972	0.063	0.172	0.949	1	1	1	1
OXY*	-2.295	5.580	-0.411	0.557	0.681	1	1	1	1
PDE	0.352	4.999	0.070	0.179	0.944	1	1	1	1
PDT	0.531	5.296	0.100	0.025	0.920	1	1	0.911	1
TNA*	-4.479	5.262	-0.851	0.731	0.395	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 9: DR-IPW estimates for associations between birthweight and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	0.040	46.165	0.001	0.999	0.999	1	1	1	1
28	14.313	47.137	0.304	0.695	0.761	1	1	1	1
44	-11.813	50.293	-0.235	0.925	0.814	1	1	1	1
49	7.162	46.031	0.156	0.945	0.876	1	1	1	1
52	-1.311	47.579	-0.028	0.984	0.978	1	1	1	1
66	-59.061	49.707	-1.188	0.160	0.235	1	1	1	1
74*	28.732	55.910	0.514	0.541	0.607	1	1	1	1
99*	-7.170	45.603	-0.157	0.880	0.875	1	1	1	1
101	-48.104	49.721	-0.967	0.171	0.333	1	1	1	1
118*	-13.949	62.067	-0.225	0.802	0.822	1	1	1	1
138*	-16.715	52.475	-0.319	0.720	0.750	1	1	1	1
146*	18.038	53.847	0.335	0.737	0.738	1	1	1	1
153*	-27.790	65.033	-0.427	0.709	0.669	1	1	1	1
156*	-7.903	43.303	-0.182	0.852	0.855	1	1	1	1
180*	87.079	63.803	1.365	0.142	0.172	1	1	1	1
183*	-54.862	50.736	-1.081	0.237	0.280	1	1	1	1
187*	-33.277	43.122	-0.772	0.355	0.440	1	1	1	1
194*	0.134	39.825	0.003	0.998	0.997	1	1	1	1
201*	41.021	36.701	1.118	0.236	0.264	1	1	1	1
BHC*	-24.662	49.647	-0.497	0.610	0.619	1	1	1	1
DIE*	7.775	55.984	0.139	0.873	0.890	1	1	1	1
GHC	-43.620	44.687	-0.976	0.351	0.329	1	1	1	1
HCB*	-28.984	41.487	-0.699	0.411	0.485	1	1	1	1
HPE*	-80.821	41.927	-1.928	0.020	0.054	1	1	0.862	0.993
MIR*	33.300	46.303	0.719	0.472	0.472	1	1	1	1
ODT	-17.191	48.316	-0.356	0.653	0.722	1	1	1	1
OXY*	-35.522	53.069	-0.669	0.484	0.503	1	1	1	1
PDE	52.237	45.836	1.140	0.290	0.254	1	1	1	1
PDT*	26.038	53.024	0.491	0.649	0.623	1	1	1	1
TNA*	-36.243	47.960	-0.756	0.398	0.450	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 10: DR-IPW estimates for associations between gestational age and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	0.138	0.140	0.987	0.128	0.324	1	1	1	1
28	0.014	0.146	0.095	0.942	0.924	1	1	1	1
44	-0.047	0.158	-0.300	0.928	0.764	1	1	1	1
49	0.046	0.141	0.324	0.624	0.746	1	1	1	1
52	0.112	0.140	0.799	0.290	0.424	1	1	1	1
66	-0.186	0.157	-1.181	0.080	0.238	1	1	1	1
74*	-0.084	0.164	-0.514	0.531	0.607	1	1	1	1
99*	0.035	0.131	0.263	0.765	0.792	1	1	1	1
101	0.028	0.161	0.173	0.807	0.862	1	1	1	1
118*	-0.169	0.181	-0.936	0.282	0.349	1	1	1	1
138*	-0.085	0.177	-0.480	0.587	0.631	1	1	1	1
146*	0.181	0.152	1.193	0.254	0.233	1	1	1	1
153*	-0.182	0.203	-0.897	0.422	0.370	1	1	1	1
156*	0.051	0.123	0.416	0.675	0.678	1	1	1	1
180*	-0.135	0.189	-0.717	0.439	0.474	1	1	1	1
183*	0.171	0.154	1.109	0.254	0.267	1	1	1	1
187*	0.031	0.177	0.177	0.836	0.859	1	1	1	1
194*	-0.027	0.167	-0.163	0.921	0.870	1	1	1	1
201*	0.069	0.162	0.423	0.656	0.673	1	1	1	1
BHC*	0.143	0.150	0.954	0.301	0.340	1	1	1	1
DIE*	0.047	0.201	0.234	0.777	0.815	1	1	1	1
GHC	0.008	0.141	0.060	0.949	0.952	1	1	1	1
HCB*	-0.003	0.169	-0.020	0.982	0.984	1	1	1	1
HPE*	0.168	0.151	1.110	0.205	0.267	1	1	1	1
MIR*	-0.096	0.142	-0.676	0.433	0.499	1	1	1	1
ODT	0.012	0.163	0.071	0.936	0.943	1	1	1	1
OXY*	0.005	0.144	0.037	0.966	0.971	1	1	1	1
PDE	0.096	0.152	0.630	0.502	0.528	1	1	1	1
PDT*	0.108	0.170	0.636	0.535	0.525	1	1	1	1
TNA*	-0.093	0.165	-0.568	0.543	0.570	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 11: DR-IPW estimates for associations between head circumference and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	0.127	0.136	0.935	0.086	0.350	1	1	1	1
28	0.127	0.140	0.908	0.249	0.364	1	1	1	1
44	0.139	0.157	0.888	0.257	0.374	1	1	1	1
49	0.094	0.142	0.661	0.311	0.509	1	1	1	1
52	0.075	0.145	0.515	0.386	0.607	1	1	1	1
66	-0.057	0.158	-0.364	0.608	0.716	1	1	1	1
74*	0.079	0.163	0.488	0.546	0.626	1	1	1	1
99*	-0.084	0.163	-0.514	0.662	0.607	1	1	1	1
101	0.070	0.157	0.445	0.590	0.656	1	1	1	1
118*	-0.095	0.197	-0.484	0.669	0.629	1	1	1	1
138*	-0.173	0.195	-0.888	0.336	0.375	1	1	1	1
146*	-0.240	0.185	-1.300	0.124	0.194	1	1	1	1
153*	-0.216	0.193	-1.116	0.240	0.265	1	1	1	1
156*	-0.311	0.169	-1.841	0.040	0.066	1	1	0.977	0.998
180*	0.140	0.218	0.642	0.480	0.521	1	1	1	1
183*	-0.262	0.182	-1.438	0.107	0.150	1	1	1	1
187*	-0.250	0.190	-1.316	0.131	0.188	1	1	1	1
194*	-0.153	0.188	-0.814	0.378	0.416	1	1	1	1
201*	-0.332	0.157	-2.112	0.010	0.035	1	1	0.651	0.962
BHC*	-0.317	0.150	-2.114	0.078	0.035	1	1	1	0.961
DIE*	-0.107	0.176	-0.607	0.496	0.544	1	1	1	1
GHC	-0.101	0.157	-0.645	0.577	0.519	1	1	1	1
HCB*	-0.339	0.178	-1.904	0.033	0.057	1	1	0.955	0.995
HPE*	-0.507	0.172	-2.938	0.001	0.003	0.096	0.396	0.079	0.285
MIR*	-0.063	0.163	-0.386	0.652	0.700	1	1	1	1
ODT	-0.150	0.167	-0.900	0.322	0.368	1	1	1	1
OXY*	-0.411	0.200	-2.059	0.020	0.039	1	1	0.859	0.974
PDE*	-0.026	0.151	-0.173	0.851	0.863	1	1	1	1
PDT	0.025	0.171	0.148	0.908	0.882	1	1	1	1
TNA*	-0.250	0.176	-1.415	0.126	0.157	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 12: DR-IPW estimates for associations between length and each exposure (exp), standard errors (SE), test statistics (T), raw p-values from conditional permutation (CPD) and standard Normal (Norm) distributions, Bonferroni (Bon) adjusted and Quantile Transformation Method (QTM) adjusted p-values from both distributions.

exp	Estimate**	SE	T	Raw CPD	Raw Norm	Bon CPD	Bon Norm	QTM CPD	QTM Norm
18	-0.351	0.225	-1.560	0.376	0.119	1	1	1	1
28	-0.197	0.236	-0.832	0.693	0.406	1	1	1	1
44	-0.270	0.270	-1	0.421	0.317	1	1	1	1
49	-0.391	0.238	-1.642	0.145	0.101	1	1	1	1
52	-0.310	0.250	-1.236	0.287	0.217	1	1	1	1
66*	-0.266	0.272	-0.978	0.228	0.328	1	1	1	1
74*	0.101	0.280	0.361	0.649	0.718	1	1	1	1
99*	0.117	0.230	0.510	0.684	0.610	1	1	1	1
101	-0.397	0.273	-1.454	0.148	0.146	1	1	1	1
118*	0.081	0.325	0.249	0.791	0.803	1	1	1	1
138*	0.315	0.314	1.004	0.320	0.316	1	1	1	1
146*	0.217	0.329	0.659	0.439	0.510	1	1	1	1
153*	-0.065	0.320	-0.202	0.848	0.840	1	1	1	1
156*	0.008	0.271	0.028	0.979	0.978	1	1	1	1
180*	0.635	0.315	2.015	0.023	0.044	1	1	0.895	0.983
183*	0.053	0.325	0.162	0.859	0.872	1	1	1	1
187*	0.224	0.275	0.812	0.387	0.417	1	1	1	1
194*	0.034	0.246	0.138	0.879	0.891	1	1	1	1
201*	0.372	0.241	1.542	0.103	0.123	1	1	1	1
BHC	0.114	0.287	0.398	0.651	0.690	1	1	1	1
DIE	-0.264	0.301	-0.877	0.350	0.381	1	1	1	1
GHC	-0.180	0.235	-0.766	0.381	0.444	1	1	1	1
HCB*	-0.321	0.266	-1.209	0.175	0.227	1	1	1	1
HPE*	-0.326	0.287	-1.135	0.213	0.256	1	1	1	1
MIR*	0.067	0.267	0.250	0.789	0.802	1	1	1	1
ODT	0.315	0.274	1.151	0.184	0.250	1	1	1	1
OXY*	-0.134	0.308	-0.435	0.631	0.664	1	1	1	1
PDE	0.352	0.237	1.486	0.137	0.137	1	1	1	1
PDT	0.531	0.283	1.874	0.023	0.061	1	1	0.897	0.997
TNA*	-0.294	0.298	-0.988	0.294	0.323	1	1	1	1

*Estimate based on at least one weight $\hat{g}(0|W_i), i = 1, \ldots, n$ truncated to 0.1
**See Tables 1 through 4 for sample sizes (N).

Table 13: Variables in W selected by the DSA for estimates of g_j and $Q_{j,k}$ for estimated associations between birthweight ($k = 1$) and each exposure (exp) ($j = 1, \ldots, 30$).

exp	g_j	$Q_{j,k}$
18	none	parity, educ
28	educ, US years	none
44	none	parity
49	none	none
52	none	parity, country
66	US years	none
74	age	educ, parity, BMI, sex
99	age, US years	none
101	US years	parity, educ
118	age, US years	none
138	age, educ	educ, parity
146	educ, age	none
153	educ, age	country, poverty
156	age, marital, educ	none
180	educ, age, BMI	none
183	educ, age, marital, BMI, US years	none
187	age, US years, BMI, poverty, marital, educ	country, poverty
194	age, US years, BMI	country, parity, BMI, educ, sex, gest age
201	age, US years, BMI	educ, parity
BHC	country, parity, educ, US years, age	country, poverty
DIE	US years, age, marital	parity, educ, country, BMI
GHC	sex, PCB sum	country, pooverty
HCB	PCB sum, age, parity, US years	educ, parity, sex, BMI, country, marital
HPE	sex, country, age, PCB sum	none
MIR	age	none
ODT	educ, US years	educ, parity, sex, BMI, country, marital
OXY	age, parity, US years, educ, PCB sum, country, BMI	none
PDE	country, US years	none
PDT	age, country	educ, parity, sex, BMI, country
TNA	PCB sum, US years, age, parity	educ, parity, sex, BMI, country, marital

Table 14: Estimated standard errors (SE) of $\hat{\psi}_{1,3}$ based on both the influence curve (IC) and the bootstrap (BS), as well as the bootstrap estimated variance (VAR) of $T = \frac{\hat{\psi}}{\text{se}(\hat{\psi}_{1,3})}$, where $\text{se}(\hat{\psi}_{1,3})$ is obtained using the influence curve for $\hat{\psi}_{1,3}$ for the association between PCB 18 and head circumference based on the DR-IPW estimator. Bootstrap estimates based on 5000 bootstrap samples of (3).

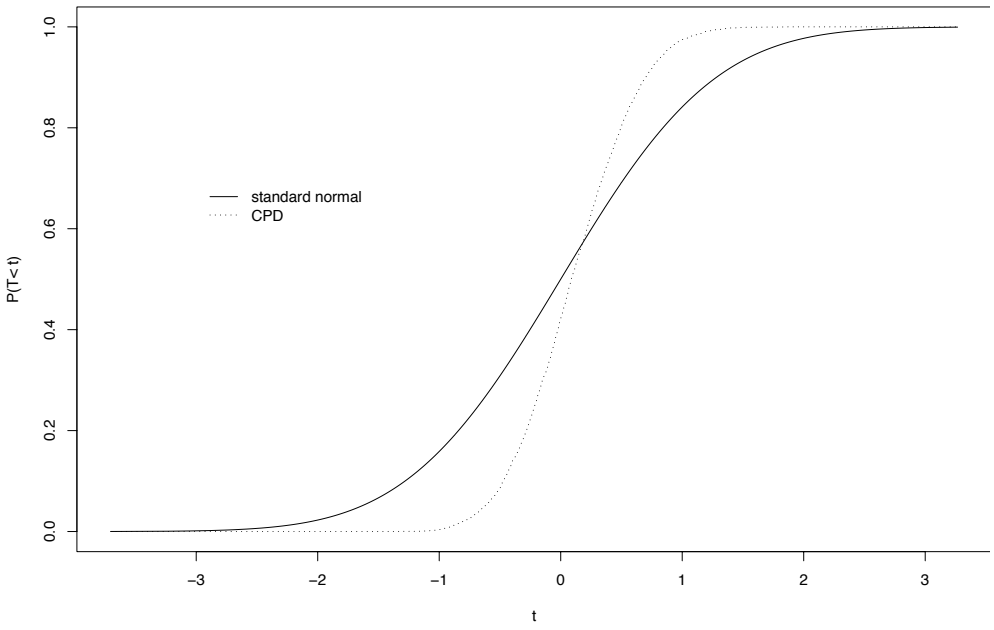
IC SE($\hat{\psi}_{1,3}$)	BS SE($\hat{\psi}_{1,3}$)	BS VAR(T)
0.136	0.116	0.704

Table 15: Overall summary of the bootstrapped (BS) estimates of the variance (VAR) of the test statistic based on the IPW estimator (T_{ipw}) and and the DR-IPW estimator (T_{dr}) .

	BS VAR(T_{ipw})	BS VAR(T_{dr})
Min	0.001	0.454
Median	0.127	0.776
Mean	0.150	0.771
Max	0.660	1.082
$> 1^*$	0	1

*Number of test statistics where bootstrapped estimate of the variance greater than one.

Figure 1: Plot of the cumulative distribution function of the standard Normal vs. that of the conditional permutation distribution (CPD) for the test statistic (T) based on the DR-IPW estimator of (3) associating head circumference and PCB 18.



As we can see from Tables 5 through 12, several tests of association which would be classified as significant or borderline significant ($p < 0.1$) based solely on raw p-values are no longer significant after application of a MTP, regardless of type (Bonferroni or Quantile Transformation Method). Based on the IPW analysis, there were 11 tests of association with raw p-values (obtained from either a standard Normal or the modified conditional permutation distribution) less than 0.1. In all cases the respective adjusted p-values were greater than 0.9. Based on the DR-IPW analysis, there were also 11 tests with raw p-values less than 0.1. In all but one case (the association between HPE and head circumference) respective adjusted p-values were greater than 0.6. Notably, the adjusted p-value for the association between HPE and head circumference based on the DR-IPW estimator remained borderline significant whether the Bonferroni or QTM approach was used. However, for this test of association, adjusted p-values based on both MTPs were substantially larger when the standard Normal was used over the conditional permutation distribution (see Table 11). This is in line with the results presented in Table 15, indicating that inference based on the standard Normal tended to be more conservative than that based on the modified conditional permutation distribution.

7 Discussion

In summary, we found only one borderline significant association in the CHAMACOS data set after analysis with the proposed machine learning algorithm; that between HPE and head circumference. Contrary to expectations, the direction of the estimate in this case suggests a protective effect of HPE. We consider two explanations. First, there is a violation of at least one of our identifying assumptions. It is certainly possible that our definition of W did not include all relevant confounders of the exposure effect, considering how little is known regarding the effects of organochlorines on human development and the possible relationships between different OC's and PCB's Fenster et al. [2007].

As expected, raw p-values were quite different from adjusted p-values indicating the importance of adjustment for multiple testing. In this application, conclusions were similar regardless of what multiple testing method was used (Bonferroni or Quantile Transformation Method). Also, as expected, results differed depending on whether a standard Normal or the modified conditional permutation distribution was assumed for the test statistics, with the standard Normal generally resulting in more conservative inference.

As mentioned in §5.2 and §5.3, due to indeterminately long computing times required for calls to the DSA algorithm with $B = B_2 = 5,000$, we did not re-select the forms of g and Q within each permutation and bootstrap iteration. Failure to re-select these model forms within each permutation and bootstrap iteration may affect results in an unpredictable direction. Simulation results suggested that this shortcut does not substantially impact results. In theory, faster data-adaptive model selection algorithms are available as alternatives to the DSA algorithm. We attempted to use the R functions *polyclass()* for selection of g_j and *polymars()* for selection of $Q_{j,k}$ but found *polyclass()* too unstable with $B = B_2 = 5,000$ when called within each iteration. It is possible that repeated calls to the DSA algorithm may be more feasible with the use of a computer cluster to decrease computation time.

We note that results based on this analysis differ from other reported analyses of the CHAMACOS data set. In a recently published report of associations between OCs and birth outcomes in the CHAMACOS data set, Fenster et al. [2007] reported a significant association between gestational age and HCB and did not find a signficant relationship between HPE and head circumference. We stress that results reported by Fenster et al. [2007] are not generally comparable to those reported here beyond differences in approaches to marginal inference and multiple testing. Specifically:

- our parameter is based on a categorical measure of exposure and represents a marginal effect over W while their parameter is based on continuous exposure measures and conditional on variables in W
- we consider a slightly different set of variables in W;

- different approaches to model selection were used;
- slightly different exclusionary criteria were used to obtain the final analysis sample.

Thus, the lack of correspondence of other analyses comes from both a different parameter being estimated as well as a different method used to estimate the parameter. Our method is most suitable when little existing information is available to choose models/variables *a priori* and thus the information for the relative contribution to variability in the outcome comes almost exclusively from the data itself. In this context, one wants a procedure that produces a simple, interpretable parameter, uses flexible semi-parametric (machine learning) methods for models when these are not known a priori (and the dimension is high) and finally returns trustworthy joint inference. We believe many studies pretend as if knowledge of the model exists or the exploring of the data for such a model is ignored in the final inference. We also believe this leads to erroneous estimates (bias due to model misspecification) as well as erroneous inference (standard errors based on the assumption of a priori known models). This method is a potential black-box tool that can be used to screen for the variables with strong evidence of adjusted associations. Traditional ad hoc approaches have served useful purposes, but the combination of new techniques in causal inference, more powerful machine learning tools, and fast computation that allows robust, re-sampling based inference means the practice of exploratory epidemiology should move beyond potentially misleading approaches appropriate for low dimensional problems and provide more robust results for high dimensional studies.

References

US Census Bureau. Poverty Thresholds 2000, Current Population Survey, 2000. Technical report, 2000. URL `http://www.census.gov/hhes/poverty/poverty00/pv00thrs.html`.

J. Chen, M.J. van der Laan, M.T. Smith, and A.E. Hubbard. A comparison of methods to control type i errors in microarray studies. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007. URL `http://www.bepress.com/sagmb/vol6/iss1/art28`.

J. Chevrier, B. Eskenazi, N. Holland, A. Bradman, and D.B. Barr. Effect of exposure to polychlorinated biphenyls and organochlorine pesticides on thyroid function during pregnancy. *American Journal of Epidemiology*, 68(3):298–310, 2008.

S. Dudoit, M.J. van der Laan, and K.S. Pollard. Multiple testing, Part i. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004. URL `http://www.bepress.com/sagmb/vol3/iss1/art13`.

B. Eskenazi, A.R. Marks, A. Brandman, L. Fenster, C. Johnson, D.B. Barr, and N.P. Jewell. In utero exposure to dichlorodiphenyltrichloroethane (ddt) and

dichlorodiphenyldichloroethylene (dde) and neurodevelopment among young mexican american children. *Pediatrics*, 118(1):233–41, 2006.

L. Fenster, B. Eskenazi, M. Anderson, A. Bradman, K. Harley, H. Hernandez, A. Hubbard, and D.B. Barr. Association of *in utero* organochlorine pesticide exposure and fetal growth and length of gestation in an agricultural population. *Environmental Health Perspectives*, 115(6), April 2007. URL `http://www.ehponline.org/docs/2005/8423/abstract.html`.

A.E. Hubbard and M. J. van der Laan. Population Intervention Models in Causal Inference. Technical Report 191, Division of Biostatistics, University of California, Berkeley, 2005. URL `http://www.bepress.com/ucbbiostat/paper191`.

K.S. Pollard and M.J. van der Laan. Resampling-based Multiple Testing: Asymptotic Control of Type I Error and Applications to Gene Expression Data. Technical Report 121, Division of Biostatistics, University of California, Berkeley, 2003. URL `http://www.bepress.com/ucbbiostat/paper121`.

J.M. Robins. Marginal structural models. In *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, pages 1–10. American Statistical Association, 1998.

P.R. Rosenbaum. Conditional permutation tests and the propensity score in observational studies. *Journal of the American Statistical Association*, 79(387):565–574, 1984.

S.E. Sinisi and M.J. van der Laan. Loss-Based Cross-Validated Deletion/Substitution/Addition Algorithms in Estimation. Technical Report 143, Division of Biostatistics, University of California, Berkeley, 2004. URL `http://www.bepress.com/ucbbiostat/paper143`.

M. J. van der Laan and A.E. Hubbard. Quantile-Function Based Null Distribution in Resampling Based Multiple Testing. Technical Report 198, Division of Biostatistics, University of California, Berkeley, 2005. URL `http://www.bepress.com/ucbbiostat/paper198`.

M. J. van der Laan, S. Dudoit, and K.S. Pollard. Multiple testing, Part III. Procedures for Control of the Generalized Family-Wise Error Rate and Proportion of False Positives. Technical Report 141, Division of Biostatistics, University of California, Berkeley, 2004a. URL `http://www.bepress.com/ucbbiostat/paper141`.

Mark J. van der Laan and Alan E. Hubbard. Quantile-function based null distribution in resampling based multiple testing. *Statistical Applications in Genetics and Molecular Biology*, 5(1), 2006. URL `http://www.bepress.com/sagmb/vol5/iss1/art14`. Article 14.

Mark J. van der Laan and Maya Petersen. Causal effect models for realistic individualized treatment and intention to treat rules. *International Journal of Biostatistics*, 3(1), 2007. URL `http://www.bepress.com/ijb/vol3/iss1/3`. Article 3.

M.J. van der Laan and J.M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer, 2003.

M.J. van der Laan, S. Dudoit, and K.S. Pollard. Multiple testing, Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004b. URL `http://www.bepress.com/sagmb/vol3/iss1/art14`.

Yue Wang, Maya L. Petersen, David Bangsberg, and Mark J. van der Laan. Diagnosing bias in the inverse probability of treatment weighted estimator resulting from violation of experimental treatment assignment. Technical Report 211, Division of BiostatWantistics, University of California, Berkeley, 2006. URL `http://www.bepress.com/ucbbiostat/paper211`.